520.42989X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):	SHIROGANE, et al.
Serial No.:	Not yet assigned
Filed:	July 30, 2003
Title:	STORAGE SYSTEM
Group:	Not yet assigned

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

July 30, 2003

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Patent Application No.(s) 2002-335301, filed November 19, 2002.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Ronald J. Shore
Registration No. 28,577

RJS/alb
Attachment
(703) 312-6600

日　本　国　特　許　庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２００２年１１月１９日

出願番号
Application Number: 特願２００２－３３５３０１

[ST.10/C]: [ＪＰ２００２－３３５３０１]

出願人
Applicant(s): 株式会社日立製作所

２００３年　６月　２日

特許庁長官
Commissioner,
Japan Patent Office 太田信一郎

出証番号　出証特２００３－３０４１８９７

【書類名】　特許願
【整理番号】　NT02P0643
【提出日】　平成14年11月19日
【あて先】　特許庁長官　殿
【国際特許分類】　H04B　7/00
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　白銀　哲也
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　上村　哲也
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　田中　淳
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社日立製作所
【代理人】
　【識別番号】　100068504
　【弁理士】
　【氏名又は名称】　小川　勝男
　【電話番号】　03-3661-0071
【選任した代理人】
　【識別番号】　100086656
　【弁理士】
　【氏名又は名称】　田中　恭助

【電話番号】　03-3661-0071

【選任した代理人】

【識別番号】　100094352

【弁理士】

【氏名又は名称】　佐々木　孝

【電話番号】　03-3661-0071

【手数料の表示】

【予納台帳番号】　081423

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　ストレージシステム

【特許請求の範囲】

【請求項１】

情報発生部からの情報をパケット化して、ネットワークに送信し、またはネットワークからパケットを受信する情報処理装置において、情報発生部からの情報をＴＣＰ／ＩＰパケット群に変換する手段と、送信する相手先に対応してＦＥＣ(Forward Error Correction)の冗長度を管理する管理部と、該管理部に保持している相手先の冗長度を参照して、ＴＣＰ／ＩＰ変換されたパケット群に対してＦＥＣエンコード処理を行うエンコード部と、ネットワークから受信されるパケット群の情報をＦＥＣデコード処理するデコード部とを有する情報処理装置。

【請求項２】

前記管理部はメモリに登録されるテーブルを有し、テーブル内に登録される冗長度は相手先ごとに変更可能である請求項１記載の情報処理装置。

【請求項３】

情報処理装置は、情報発生部にディスクドライブを有するストレージ装置であり、かつ情報発生部からの情報をＴＣＰ／ＩＰパケットに変換する手段はiSCSIプロトコル処理手段であり、エンコード部はiSCSIのパケット群をエンコードし、デコード部はネットワークから受診したパケット群をデコード処理してiSCSIのパケット群とする請求項１の情報処理装置。

【請求項４】

情報処理装置において、該エンコード部でＦＥＣエンコード処理されたデータはＵＤＰパケット群としてネットワークに送信され、かつネットワークから受信されたＵＤＰパケット群は該デコード部でＦＥＣデコード処理される請求項１の情報処理装置。

【請求項５】

ネットワーク側のポートとストレージ装置側のポートとの間に介在し、パケットデータの送受信を中継する装置であって、

送信先に対応してＦＥＣ冗長度を登録して管理する送信管理テーブルと、受信先

に対応してＦＥＣ冗長度を登録して管理する受信管理テーブルと、ストレージ装置で生成されるパケット化されたiSCSI層のデータを、該送信管理テーブルを参照して送信先に応じた冗長度を持たせてＦＥＣエンコード処理するエンコード部と、該受信管理テーブルを参照して、ネットワークから受信されるパケットデータをＦＥＣデコード処理し、iSCSI層のデータに復号化するデコード部とを有する中継装置。

【請求項６】

前記送信管理テーブルはＦＥＣ処理が可能な送信先アドレスを登録し、かつ前記受信管理テーブルはＦＥＣ処理が可能な送信元アドレスを登録しており、該送信管理テーブルを参照して送信先アドレスが登録されていれば、iSCSIデータは前記エンコード部でＦＥＣ処理されてネットワークへ送信され、送信先アドレスが登録されていなければ、iSCSIデータは該エンコード部でＦＥＣ処理されずにネットワークへ送信され、一方ネットワークから受信されるパケットデータに対し、該受信管理テーブルを参照して送信元アドレスが登録されていれば該デコード部でＦＥＣデコード処理してiSCSIデータに復号化し、登録されていなければ該デコード部でＦＥＣ復号化されずにiSCSI層へ送る請求項5記載の装置。

【請求項７】

ネットワークを介して受信される制御フレームの内容を解析してアドレスを追加または削除するようにして、前記送信管理テーブルの内容及び前記受信管理テーブルの内容を変更する手段を有する請求項5記載の装置。

【請求項８】

ｉＳＣＳＩプロトコルを用いた装置間でネットワークを介してデータを伝送する通信方法であって、ＦＥＣ通信モードでデータの送受信を行う第１の通信モードと、ＴＣＰ／ＩＰ通信モードでデータの送受信を行う第２の通信モードと、データの通信先となる相手のiSCSI ネームをメモリに登録して管理するステップと、送信先に対応してＦＥＣの冗長度をメモリに登録して管理するステップと、データを送信する相手先のiSCSI ネームがメモリに登録されているか否かをチェックするステップと、チェックの結果、iSCSI ネームがメモリに登録されていればメモリに登録されているＦＥＣの冗長度に基づいてデータをＦＥＣ処理して該第１

の通信モードに従って送信するステップと、チェックの結果、iSCSI ネームがメモリに登録されていなければ第２のモードでデータを送信するステップとを有する通信方法。

【請求項９】

請求項８において、更に通信先毎に送信するパケットの喪失率を求めて管理するステップと、パケットの喪失率に従って通信先毎に該メモリに登録されている冗長度を変更するステップとを有する通信方法。

【請求項１０】

請求項８において、更に受信側の装置で、送信元に対応してＦＥＣの冗長度をメモリに登録して管理するステップと、データを受信するとき、送信元のiSCSI ネームが該メモリに登録されているか否かをチェックするステップと、チェックの結果、iSCSI ネームがメモリに登録されていればメモリに登録されているＦＥＣの冗長度に基づいてデータをiSCSIデータに復元処理するステップとを有する通信方法。

【請求項１１】

請求項８において、受信側の装置で、iSCSIデータに復元された場合にはＡＣＫを送信元に返送し、復元されなかった場合にはＡＣＫを送出しないようにするステップと、送信元の装置ではＡＣＫを受け取らなかった場合に、先に送信した同じデータを同じ送信先にＦＥＣ処理して該第１の通信モードに従って送信するステップとを有する通信方法。

【請求項１２】

複数のストレージ装置がネットワークに接続されてデータの送受信を行うストレージシステムにおいて、該ストレージ装置はデータを記録するディスクドライブと、該ディスクドライブに接続されるディスクアダプタと、該ディスクアダプタと接続されるキャッシュメモリと、該キャッシュメモリに接続されるチャネルアダプタと、該ディスクドライブからのデータをＴＣＰ／ＩＰ上のｉＳＣＳＩパケット群に変換する手段と、送信する相手先に対応してＦＥＣ(Forward Error Correction)の冗長度を管理する管理部と、該管理部に保持している相手先の冗長度を参照して、変換されたＴＣＰ／ＩＰ上のｉＳＣＳＩパケット群に対してＦＥ

Ｃエンコード処理を行うエンコード部と、ネットワークから受信されるパケット群の情報をＦＥＣデコード処理するデコード部とを有するストレージシステム。

【請求項１３】

データを記録して処理するアプリケーション層と、該アプリケション層のデータに対してＳＣＳＩ変換するｉＳＣＳＩ層と、ｉＳＣＳＩ層のデータをＴＣＰ/ＩＰ処理するＴＣＰ層及びＩＰ層を有するストレージ装置よりネットワークに対してデータの送受信を行うストレージシステムにおいて、ｉＳＣＳＩ層からのデータに対して、冗長符号を付加するＦＥＣエンコード処理層と、エンコード処理されたデータにＵＤＰ処理するＵＤＰ層と、ＵＤＰ層からのデータに対してＩＰ処理を行うＩＰ層とを有し、該ＩＰ層で処理されたデータをネットワークに送出することを特徴とするストレージシステム。

【請求項１４】

請求項１３において、更に、ネットワークから受信されたデータであって、該ＩＰ層でＩＰ処理され、ＵＤＰ層でＵＤＰ処理されたデータをＦＥＣデコード処理するＦＥＣデコード層を有するストレージシステム。

【請求項１５】

請求項１３において、該ＦＥＣエンコード処理層における冗長符号の冗長度をデータの送信先毎に変更する手段を有するストレージシステム。

【請求項１６】

ｉＳＣＳＩプロトコルを用いた装置のｉＳＣＳＩ間でネットワークを介してデータを伝送する通信方法であって、ＦＥＣ処理した通信モードでデータの送受信を行う第１の通信モードと、ＴＣＰ／ＩＰ通信モードでデータの送受信を行う第２の通信モードと、データの通信先となる相手がｉＳＣＳＩ層を有するか否かをiSCSI ネームを用いてチェックするステップと、このチェックの結果、相手がｉＳＣＳ層を有する場合、相手に応じたＦＥＣの冗長度に基づいてデータをＦＥＣ処理して該第１の通信モードに従って送信するステップと、チェックの結果、相手がiSCSI層を有しない場合、第２のモードでデータを送信するステップとを有する通信方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明はストレージシステムに係り、特にネットワークを介して接続されるＲＡＩＤのようなストレージ装置間、又はストレージ装置とホスト間のデータ通信方法、及びそれを実現させるためのストレージ装置の構成に関するものである。

【０００２】

【従来の技術】

ＩＰストレージの普及及びデータのリモートコピーに対するニーズの増大により、ネットワークを介して接続された遠距離にあるストレージ間或いはストレージ装置とサーバ間でデータ転送することが頻繁になってきている。しかも、転送するデータ量は年々増加しており、大容量のデータを高速、高信頼で送信することが必要となってきている。

【０００３】

サーバとストレージ装置間を接続するとインタフェースして、ファイバチャネルやＳＣＳＩインタフェースが知られている。ファイバチャネルはＳＡＮ(Storage Area Network)を構成する高速データ転送のための標準的なインタフェースとして使用されている。また、ＳＣＳＩインタフェースは、データ伝送速度が速く、伝送遅延も小さい上に、伝送エラーの発生確率が非常に低いと言うメリットがあるので、従来からストレージ装置用のプロトコルとして良く使用されているが、ＳＣＳＩインタフェースは伝送距離が短いと言う問題がある。

【０００４】

最近、iSCSIプロトコルを用いてストレージ装置を接続する試みが行われている。iSCSIとは、ネットワーク技術であるＴＣＰ／ＩＰ上で、インタフェース技術としてＳＣＳＩ処理を実現するためのプロトコル技術であり、現在ＩＥＴＦ(the Internet Engineering Task Force)で規格化が進められている。このiSCSIプロトコルにより長距離にあるストレージ装置間でデータ転送する場合、ＩＰ層でパケットの喪失が多発することが予想される。

【０００５】

一般にデータ通信におけるパケットの喪失の対策として、ＡＲＱ(Automatic R

epeat reQest)方式即ち自動再送要求方式や、ＦＥＣ(Forward Error Correction)方式即ち前方誤り訂正方式が知られている。長距離のデータ伝送を行う場合、パケットの喪失が頻繁に生じる可能性がある。そのため、ＡＲＱ方式のデータ通信によれば、データ再送のための時間がかかり、結果的にデータの伝送効率が低下する。特に、長距離データ伝送ではデータ再送にかかる時間のうち、ネットワーク遅延が距離に従って増大するため、ＡＲＱ方式は好ましいとは言えない。

【０００６】

例えば、特開２００１－７７８５号公報(特許文献１)には、ＦＥＣ方式とＡＲＱ方式を併用し、伝送路の回線品質に適した誤り制御方式を通信相手に通知するためにユニークワードを付加して送信し、受信側ではユニークワードに応じて複合化する技術が開示されている。また、特開２００１－１６８９４４号公報（特許文献２）には、属性の異なる２種以上のデータ（即ちＩＳＯデータとＡＳＹデータ）を送信し、受信側ではこれらの属性に応じて受信処理する方式が提唱されている。

【０００７】

【特許文献１】

特開２００１－７７８５号公報

【特許文献２】

特開２００１－１６８９４４号公報

【０００８】

【発明が解決しようとする課題】

然るに、上記文献１及び文献２には、ストレージ装置との関係が開示されていないし、またＦＥＣ冗長度の変化にどのように対応するのかについても開示がない。

【０００９】

本発明の目的は、ネットワークを介して行うデータ通信においてパケットが喪失した場合でも送信データの復元が可能な通信方法を提供することにある。

本発明の他の目的は、FEC方式を採用したiSCSIプロトコル準拠のストレージ装置を提供することにある。

本発明の他の目的は、iSCSI層間でデータの伝送を行う場合、データの送信先毎にFEC処理の状態やデータ送信の冗長度を変えてデータ送信し、復元することができるストレージシステムを提供することにある。

【００１０】

【課題を解決するための手段】

上記目的を達成するために、本発明は、ストレージ装置どうし又はストレージ装置とホストコンピュータ（一般的な意味のコンピュータであり、サーバも含む）とがネットワークを介して接続されるストレージシステムにおいて、ストレージ装置はiSCSIプロトコルを採用し、iSCSI層間どうしで夫々データの送受信を行うシステムを達成する。ストレージ装置はFEC制御ユニットを有し、送信側ではこのFECユニットでデータを冗長化(エンコード処理)し、この冗長化したパケット群をネットワークに送信する。一方受信側ではネットワークから受信した冗長パケット群を元にＦＥＣユニットで受信データを復元(デコード処理)する。このようにFEC冗長化処理及び復号化処理を行うことにより、冗長化したパケット群の一部が喪失して受信できなくても、元のパケット群は復元可能である。一方受信側でデータの復元できなかった場合には、送信側にＡＣＫが返送されないので、送信側では例えばタイムアウト監視を行うことによりデータを再送することができる。

【００１１】

本発明は送信するパケットに対するFECの冗長度を変化させることができる。通信する相手のパケット喪失の状態に従って送信側でのFECの冗長度の状態を変える。冗長度の変更をする手法としては、例えば送受信するデータに対して冗長化を行ったデータ量の比率を変える等のやり方がある。また、パケットの喪失率は、例えば受信側で受信するパケット群の喪失数を送信相手毎に採取して求めることができる。これに基づいて送信側ではデータの冗長度を変えて送信する。例えば、パケットの喪失率が高ければ、FECの冗長度を高くして送信し、一方喪失率が低ければ、冗長度を下げて送信する。ネットワークによってパケットの喪失率が時々変化することが予想されるが、それに応じて冗長度を変えられる。また、パケットの喪失率に従って、パケット送信の間隔を変えることも可能であり、広

義にはパケット喪失率に従ってデータ送信の状態を変えるようにできる。
本発明は、好ましい例においては送信先に対応してＦＥＣ冗長度を登録して管理する送信管理テーブルと、受信先に対応してＦＥＣ冗長度を登録して管理する受信管理テーブルと、ストレージ装置で生成されるパケット化されたiSCSI層のデータを、送信管理テーブルを参照して送信先に応じた冗長度を持たせてＦＥＣエンコード処理するエンコード部と、受信管理テーブルを参照して、ネットワークから受信されるパケットデータをＦＥＣデコード処理し、iSCSI層のデータに復号化するデコード部とを有して構成される。これは好ましい例では、ストレージ装置に接続されるアダプタとして構成されるが、ストレージ装置内に組み込まれて構成されるようにしても良い。

【００１２】

本発明はまた、ｉＳＣＳＩプロトコルを用いたストレージシステムのｉＳＣＳＩ間でネットワークを介してデータを伝送する通信方法を実現できる。ＦＥＣ処理した通信モードでデータの送受信を行う第１の通信モードと、ＴＣＰ／ＩＰ通信モードでデータの送受信を行う第２の通信モードと、データの通信先となる相手がｉＳＣＳＩ層を有するか否かをiSCSI ネームを用いてチェックするステップと、このチェックの結果、相手がｉＳＣＳ層を有する場合、相手に応じたＦＥＣの冗長度に基づいてデータをＦＥＣ処理して該第１の通信モードに従って送信するステップと、チェックの結果、相手がiSCSI層を有しない場合、第２のモードでデータを送信するステップとを有する。
iSCSI層間で通信するストレージシステムの好ましい例によれば、TCPセッションを開始するiSCSIのログインフレームの送信を監視し、iSCSI ネームを取り出し、それが予め決められた通信相手（iSCSIノード、例えばiSCSIイニシエータやiSCSIターゲット）へのセッションである場合、このセッションが存在する間、その目的アドレスへ送信するときは冗長度を上げるように変化させて送信する。

【００１３】

【発明の実施の形態】

以下、図面を参照して本発明の実施例を説明する。
図１は本発明の一実施例によるネットワークに接続されたストレージシステムを

示すブロック図である。図１において、ＩＰネットワーク４００にはストレージ装置100，200、及びホストコンピュータ(単にホストと言う)300が接続されている。このシステムにおいて、ネットワーク400を介してストレージ装置100,200間、又はストレージ装置100，200とホスト300間でパケット形式のデータの送受信が行われる。ストレージ装置間の通信は例えばリモートコピーを行う場合であり、ストレージ装置とホスト間の通信は例えばデータ処理時、或いはデータステーションとして使用される場合である。特徴的なことはストレージ装置100，200及びホスト300は夫々ネットワークとの接続口にＦＥＣ変換アダプタ110,210,310を具備することである。

【００１４】

各ストレージ装置100，200は、通常多数のディスクドライブ101、201と、ディスクドライブ101，201にＳＣＳＩインタフェース102，202を介して接続されるディスクアダプタ103、203と、これらのディスクアダプタ103，203にバス104，204を介して接続されるキャッシュメモリ105，205と、キャッシュメモリ105，205にバス106，206を介して接続されるチャネルアダプタ107，207を具備して構成される。チャネルアダプタ107，207は夫々ポート108，208からギガビットイーサネット（登録商標）のようなな高速ＩＰインタフェース109,209を介してＦＥＣ変換アダプタ(Forward Error Correction)110,210に接続され、これらＦＥＣ変換アダプタ110，210は高速インタフェース111，211を介してＩＰネットワーク400に接続されている。チャネルアダプタ１０７，２０７は夫々ｉＳＣＳＩプロトコルの処理を行う。尚、図示していないが、高速ＩＰインタフェース111,211はポートと介してＩＰネットワーク400に接続される。

【００１５】

また、ホスト300は、情報発生部としての情報処理装置301は演算処理装置又はメモリを有し、これらは内部バス302を介してホストバスアダプタ(HBA)303に接続されている。HBA303はポート304から高速ＩＰインタフェース305を介してＦＥＣ変換アダプタ310に接続されている。ＦＥＣ変換アダプタ310はＩＰインタフェース311を介してＩＰネットワークに接続される。ホスト300では、iSCSI ＨＢＡ303がホストの入出力処理を行い、また、ホストのＯＳ(Operating System)から

指示を受けてコマンドを発行する。
上述のストレージ装置100，200およびホスト300に接続されるＦＥＣ変換アダプタ110,210,310でＦＥＣ符号のエンコード、及びデコードが行われる。このアダプタの詳細な構成は後述する。
尚、図１には示していないが、ＩＰネットワーク400には本発明の対象となるＦＥＣ変換アダプタ110,210,310を備えていない従来のストレージ装置やホストが接続されていることがある。

【００１６】

図２は本発明の実施例におけるストレージ装置に接続されるＦＥＣ変換アダプタの内部構成を示す図である。
例えばＦＥＣ変換アダプタ110は、送信系と受信系の双方の系列から構成される。即ち送信系は、ストレージ装置100側の物理層1101、ストレージ装置からのパケットデータを一時格納するバッファ１１０２、送信パケットデータをＦＥＣエンコード処理するＦＥＣエンコード部113、ＦＥＣ処理されたデータを一時格納する送信バッファ1103,及び物理層1104から構成され、物理層1104は高速ＩＰインタフェースを介してネットワークに接続される。

【００１７】

一方、受信系は、物理層1114、受信データを一時格納するバッファ1113、ＦＥＣデコード部115、バッファ1112、及び物理層1111から構成される。ＦＥＣデコード部115では、受信されたパケットデータをデコード処理してＦＥＣ復号化する。

【００１８】

またＦＥＣ変換アダプタ110は、制御プロセッサ116と、ＦＥＣ送信管理テーブル117、及びＦＥＣ受信管理テーブル118と、ＦＥＣ許可テーブル119を有している。制御プロセッサ117はアダプタ110内の全体の制御を行う。送信管理テーブル117はこのストレージ装置100から送信可能な相手先のストレージ装置またはホスト等のアドレスが登録されている。ＦＥＣ許可テーブル119はＦＥＣ通信を許容するか否かを管理するテーブルであり、通信相手を特定するコードを登録する。
尚、これらテーブルの構成は図７を参照して後述される。

【００１９】

送信管理テーブル117に送信先アドレスが登録されていれば、送信データはＦＥＣエンコード部113でＦＥＣ処理されてネットワークへ送信され、一方送信先アドレスが登録されていなければ、データはＦＥＣエンコード部113でＦＥＣ処理されずにネットワークへ送信される。送信データとなるiSCSIデータのＦＥＣ処理は、このテーブルに登録されている送信先毎の冗長度に応じて行われる。また冗長符号としては、例えば公知のＸＯＲ演算によるパリティビット符号又はReed Solomon符号等を用いることができる。

【００２０】

受信管理テーブル118は送信元となるストレージ装置等のアドレスが登録されている。この受信管理テーブル118に送信元アドレスが登録されていればＦＥＣデコード部115でＦＥＣデコード処理して内部データであるiSCSIデータに復号化し、登録されていなければデコード部115でＦＥＣ復号化されずにiSCSI層へ送られる。

ＦＥＣ許可テーブル119は通信相手を制御するための台帳としてのテーブルである。これには通信相手のiSCSIネーム毎に冗長度が登録される。

【００２１】

図３は本発明の他の実施例を示すストレージシステムのブロック図である。この例では、ストレージ装置100',及び200'共に、ディスクドライブ101,201、ディスクアダプタ103,203、キャッシュメモリ105,205、チャネルアダプタ107,207に至るまで信頼性を考慮して全て二重化されている。これに合せて、ＦＥＣ変換アダプタ110,210も夫々二重化されて、ネットワーク400に接続されている。これら各々のＦＥＣ変換アダプタ110,210がエンコード機能とデコード機能を具備している。この例では、高速インタフェース２２０にはＦＥＣ変換アダプタが接続されていない従来のポートと考えてよい。

【００２２】

更に他の例として、ネットワーク500には管理サーバ500が接続されても良い。管理サーバ500は、各ＦＥＣ変換アダプタ110,210,310が持つＦＥＣ送信管理テーブル117,ＦＥＣ受信管理テーブル118,及び許可テーブル119の管理を行うもので

あり、各通信先（例えばポート）の追加、削除を指示することにより、各ＦＥＣ変換アダプタ間の通信にＦＥＣを使用するか否かを制御する。管理テーブルはＦＥＣの通信を行うＦＥＣ変換アダプタのアドレス、またはポートのアドレスを登録することによりその管理を行う。勿論、ＦＥＣ変換アダプタを備えない従来型のストレージ装置やホストは対象とするアドレスがこのテーブルに登録されないので、ＦＥＣの効果を享有できない。

【００２３】

図４は本発明の一実施例によるデータ転送の概念を示す図である。ストレージ装置100とストレージ装置200がルータＡ，Ｂ及びネットワークを介して遠距離でデータを転送する場合を示している。ここで、ストレージ装置200はそれに限らず、ＦＥＣ変換アダプタを備えるホストまたはサーバであっても良い。

【００２４】

ストレージ装置100のアプリケーション層で処理されたデータはiSCSI層でプロトコル変換されて、更にＴＣＰ層、ＩＰ層での制御情報を付加してパケット処理されて、ネットワークインタフェース(IF)を介してＦＥＣ変換アダプタ110に送られる。

ＦＥＣ変換アダプタ110では、送信データに対してエラー訂正のための冗長符号を付加するＦＥＣ処理（エンコード処理）する。このエンコード処理については後で詳述される。エンコード処理の後ＵＤＰヘッダのポート番号を付加し、更にＩＰパケット処理して、ネットワークＩＦを介してルータＡに送る。データはルータＡのＩＰ層、ネットワークＩＦを介してネットワークに送出される。

【００２５】

一方、受信側でデータはルータＢのネットワークＩＦ、ＩＰ層を介して受信され、ＦＥＣ変換アダプタ210へ送られる。受信データはネットワークＩＦ、ＩＰ層、ＵＤＰ層を介してＦＥＣ層に送られ、そこでＦＥＣの復号化処理（デコード処理）が行われる。復号化処理においては後で詳しく説明されるが、ＦＥＣの冗長符号で訂正処理される。エラーの度合いが大きい場合にはＡＣＫが中々送信側に返送されない。そこで、送信側ではＡＣＫ受領のタイムアウトを監視することにより、もし既定時間内にＡＣＫを受信しなければ送信側から同じデータを再

送する。

復号化されたデータはストレージ装置200に送られ、同様に各層を介してiSCSI層、アプリケーション層へと送られ、使用に供される。

【００２６】

図５は通信に用いられるパケットのフォーマットの一例を示す図である。iSCSI層においては、データの通信の単位となるiSCSI PDU(Protocol Data Unit)は、ＢＨＳ(Basic Header Segment)とデータセグメントから成る。尚、ＢＨＳとデータセグメントとの間にＡＨＳ(Additional Header Sequence)が入る場合もあるが、図示の例では省略してある。ＢＨＳはメッセージの長さを格納しており、データセグメントの開始位置やメッセージの境界が分かる。iSCSI層において、イニシエータとターゲットはiSCSI PDUと言うメッセージで通信を行う。iSCSI PDUの長さは４バイトの倍数である。

ＴＣＰ層、ＩＰ層、データリンク層では上記のiSCSI層からのパケットデータに対して先頭に、データリンクヘッダ(ＤＬＨ)、ＩＰヘッダ(ＩＰＨ)，ＴＣＰヘッダ(ＴＣＰＨ)が付加される。一方iSCSIパケットデータの最後部にはデータリンクトレイラ(ＤＬＴ)が付加される。尚、イーサネットの場合にデータリンクヘッダはイーサヘッダになる。

【００２７】

図２のバッファ（エンコードバッファ）1102にはiSCSIのパケットデータに各ヘッダが付加されたものが順次格納される。そして、ＩＰヘッダ、ＴＣＰヘッダ等も含めて冗長化される。冗長化されたデータには、データリンクヘッダ(ＤＬＨ)、ＩＰヘッダ(ＩＰＨ)，ＴＣＰヘッダ(ＴＣＰＨ)の他に、更にＵＤＰヘッダ、ＦＥＣヘッダが付加されて、送信される。

【００２８】

本発明においては、ＦＥＣ通信を行う場合ＩＰアドレスではなく、ＷＷＮ(world wide name)即ちiSCSIネームをＦＥＣ許可テーブルに登録し、それによって通信先の指示等を行う。iSCSIログインを監視してＦＥＣ許可テーブルに登録されている通信先とのログインが行われたら、前述したＦＥＣ送信管理テーブル及びＦＥＣ受信管理テーブルに通信先のアドレスを登録することによりＦＥＣ通信を

開始する。ＩＰアドレスは例えばポート毎に付与されるものであるが、iSCSIネームは物理的な１つのストレージ装置に複数のネームが設定されることがある。例えば、１つのストレージを複数のパーティションに分割して複数のディスクとして使用する場合には、１つの物理的なポートに対して複数のiSCSIネームが設定される。この利用の仕方はストレージ装置の利用者にとっては大変有意義である。

このように、通信先について例えばＩＰアドレス毎ではなく、iSCSIネーム毎に管理を行う理由は、iSCSIネームの方が例えば、

“iqn.1993-11.com.disk-vender.diskarrys.sn.45678”のように、ストレージ装置の利用者にとって覚え易く、意味を持たせ易いこと、及びＩＰアドレスを管理する場合にはネットワークＩＦカード（又はネットワークアダプタカード）を変換したり、接続先のネットワークの設定が変更になったとき、ＩＰアドレスを設定し直さねばならないが、iSCSIネームではその必要が無いこと、及びＩＰアドレスはiSCSI ネームに比べ詐称される可能性が高いこと等が上げられる。

【００２９】

図６はＦＥＣヘッダのフォーマットを示す図である。

ＦＥＣヘッダは３２ビット、４ワードの構成であり、ワード０はエンコード情報、ＦＥＣ情報の種類、及びＦＥＣ制御情報ＩＤからなる。制御情報の種類により異なるワード１、２，３のＦＥＣ制御情報の内容が変る。ここで、エンコード情報は、ＦＥＣデータ部がＦＥＣ冗長化されているか否かを示す。ＦＥＣ情報の種類は、ＦＥＣ許可テーブルの情報の変更、制御テーブル情報の変更、ＦＥＣ_ＡＣＫ、パケット喪失率の報告、データ、冗長データの種類を示す。ＦＥＣ制御情報ＩＤはＦＥＣ_ＡＣＫ，ＦＥＣ_ＲＪＴが対応する制御情報を示すために用いる。ＦＥＣ制御情報には、データ長、冗長度、パケット到達率、等の情報が含まれる。ＦＥＣデータ部はＦＥＣ制御パケットにiSCSIネームが含まれる場合、データパケット、冗長パケットに存在する。

【００３０】

図７はＦＥＣの制御テーブルと送信及び受信管理テーブルの内容の一例を示す図である。

【００３１】

ＦＥＣ許可テーブル119はＦＥＣアダプタ毎に備えられており、通信相手がＦＥＣ通信の対象か否かを管理するための所謂台帳として機能する。そのためこれには通信相手毎のiSCSIネームと冗長度を登録する。このテーブルに登録されているiSCSIネームに対するログインが行われたら、ＦＥＣ送信管理テーブル又はＦＥＣ受信管理テーブルに登録される。尚、この許可テーブル119に登録されていない相手に対しては通常のＴＣＰ／ＩＰ通信を行うようにして良い。

【００３２】

ＦＥＣ送信管理テーブル117には、個々の送信先アドレス（デスティネーションアドレス）と、それに対応した冗長度と、エンコード用バッファ制御情報が登録されている。ここで、デスティネーションアドレスとしては例えばＩＰアドレスが使用される。送信管理テーブル117にデスティネーションアドレスが登録されていれば、送信データであるiSCSIデータはＦＥＣエンコード部113でＦＥＣ処理されてネットワークへ送信される。即ちＦＥＣ通信モードで送信される。一方デスティネーションアドレスが登録されていなければ、iSCSIデータはＦＥＣエンコード部113でＦＥＣ処理されずに、ＴＣＰ／ＩＰの通信モードでネットワークへ送信される。

ＦＥＣ受信管理テーブル118には、ソースアドレスに対応してパケット到達率、ＦＥＣデコード用バッファ制御情報が登録される。ソースアドレスとしては受信先のＩＰアドレスが使用される。このパケット到達率からパケット喪失率が求められれ、最終的には送信先への冗長度の変更のために反映される。

尚、これらＦＥＣ送信管理テーブル及び受信管理テーブルには、更にiSCSIネームの欄が追加されても良い。その欄にはiSCSIネームが登録され、ＦＥＣ許可テーブルからこのiSCSIネームを参照することにより、これらのエンコード、デコード管理サーブルからの登録情報の追加，削除が行われる。

【００３３】

次に図８乃至図１０を参照して、ＦＥＣ変換アダプタにおけるデータを送信するときのＦＥＣ冗長化処理の動作について説明する。

【００３４】

（Ａ）エンコード処理

まず、図８に示すフローチャートを参照してエンコード処理について説明する。ストレージ装置で生成されたデータは、iSCSIパケット群として高速ＩＰインターフェースを介してチャネルアダプタ107からＦＥＣ変換アダプタ110に送られる。ＦＥＣ変換アダプタ110では物理層1101を介してこのパケット群が受信され（801）、バッファ1102に一時的に格納される。そして、ＦＥＣエンコード部113では、パケットがiSCSI Login request PDUかまたはiSCSI Login response PDUかをチェックする（802）。その結果、否ならば、次にＦＥＣ送信管理テーブル117を参照し、デスティネーションアドレスがこの管理テーブル117登録されているか否かがチェックされる（803）。もし、登録されていれば、ＦＥＣエンコード用バッファ1103が割り当てらえる（804）。一方、制御テーブル117に登録されていなければ、パケットはＦＥＣ処理されずそのままの形式（即ちＴＣＰ／ＩＰ形式）のパケット群データとしてバッファ1103を介して物理層1104に送られ、ネットワークに送出される（816）。また、上記ステップ８０４で、ＦＥＣエンコードバッファの領域が割り当てられていなければ、新しいＦＥＣシーケンスの開始準備を行う（817）。

【００３５】

上記したＦＥＣ用エンコードバッファが割り当てられる場合、パケットを格納するアドレスが計算され（805）、パケットはそのアドレスへそのまま格納される806）。即ちＩＰパケットは丸ごとカプセル化され、ＩＰヘッダ、ＴＣＰヘッダ、ＴＣＰデータはそのままバッファに格納される。次に、パケットにはＩＰヘッダ、ＵＤＰヘッダ、ＦＥＣヘッダが付加され、カプセル化して、物理層1104を介してネットワークへ送出される（807）。ＦＥＣヘッダの各領域には対応する値を格納する。即ちＦＥＣヘッダ種にはＦＥＣデータを、データ長にはカプセル化したパケットの長さを格納し、その他はＦＥＣ送信管理テーブルの内容に従う。

その後、ＦＥＣ送信管理テーブル117の内容が更新される（808）。このテーブル117の更新はこのＦＥＣシーケンスについて格納されるパケット数をプラス“1”する。送信するパケットの総数を管理するためにある。

【００３６】

次に、ＦＥＣシーケンス（冗長度ｎ）に対してｎ個目のパケットか否かが判断され（809）、ｎ個目でなければ終了する。一方、ｎ個目であれば、最後に冗長データをn+1個目のパケットして送信するため、冗長度ｎに従ってこのＦＥＣエンコード用バッファのデータから冗長データを作成する（810）。そして、冗長データにＩＰヘッダ、ＵＤＰヘッダ、ＦＥＣヘッダを付加し、カプセル化して物理層1104を介してネットワークに送出される（811）。そしてこの一連のＦＥＣ処理シーケンスの終了処理をする（812）。終了処理とは、ＦＥＣ再送機能が有る場合には、このＦＥＣシーケンスはＦＥＣ＿ＡＣＫ待ちとするものであり、ＦＥＣ再送機能が無い場合には、ＦＥＣエンコードバッファを開放するなどの処理である。

さて、上記ステップ８０２でのチェックの結果、Yesならば、そのパケット群がiSCSI Initial Logon request PDUか否かのチェックが行われる（813）。その結果Yesであれば、ターゲットネームがＦＥＣ許可テーブルに登録されているか否かがチェックされる（814）。このテーブル登録のチェックの結果、登録されていれば、パケットのデスティネーションアドレスについてＦＥＣ送信管理テーブルに追加して次の処理へと移る（815）。

一方、上記ステップ８１３のチェックの結果、否の場合にはそのパケットをＦＥＣ処理せずそのまま物理層1104へ送り（816）、処理を終了する。

【００３７】

（Ｂ）送信データが不足している場合の処理動作

次に、図９に示すフローチャートを参照して、送信データが不足している場合の処理動作について説明する。この処理はストレージ装置のチャネルアダプタから正規の数のパケットを受け取っているか否かを確認するための処理である。

ある送信シーケンスにかかるパケット群の送信を開始すると、タイマ割り込みがかかり（820）、タイマのカウントが開始される。そしてタイマが、送信処理中のＦＥＣシーケンスの処理開始から一定時間以上経過したか否かがチェックされる（821）。経過していなければ処理は終了する。

一方、一定時間を経過していれば、このＦＥＣシーケンスについて既に物理層11

01から受信しているｍ個のパケットに対して、このＦＥＣシーケンスの冗長度をｍとして冗長データを作成する（822）。尚、この冗長度ｍは冗長パケットのＦＥＣヘッダに格納される。

【００３８】

（Ｃ）ＦＥＣパケットの再送動作

次に、図１０のフローチャートを参照してＦＥＣパケットの再送動作について説明する。

この処理の基本的シーケンスは、受信側ではＦＥＣパケットを受信するとＦＥＣ＿ＡＣＫを送信側に返送する。送信側ではこのＦＥＣ＿ＡＣＫを受領することにより、送信したパケットが正しく相手に送られたと認識する。

さて、送信側ではＦＥＣパケットを送信すると、タイマ割り込み処理が行われる（830）。そしてあるＦＥＣ＿ＡＣＫ待ち中のＦＥＣシーケンスに関してタイマがタイムアウトをしたか否かがチェックされ（831）る。このチェックの結果タイムアウトしていなければ終了する。しかし、チェック831の結果、タイムアウトした場合には、このＦＥＣシーケンスについての全パケットを再送信するための処理が行われる。同様にして、これら再送のパケットに関してＦＥＣ＿ＡＣＫを待つことになるので、タイマも再設定される（832）。

【００３９】

次に、図１１及び図１２を参照してＦＥＣ変換アダプタ110でデータを受信するときのＦＥＣデコード処理の動作について説明する。

図１１及び図１２に示すフローチャートにおいて、物理層1114からパケットを受信すると（901）、パケットがiSCSI Login request PDU 又はiSCSI Login response PDUであるか否かがチェックされる（902）。Yesであれば、そのパケットがiSCSI initial Login request PDUか否かがチェックされ（919）、その結果該当すれば、source ネーム(即ちiSCSI ネーム)がＦＥＣ許可テーブルに登録されているか否かがチェックされる（920）。許可テーブルに登録されていれば、そのパケットのデスティネーションアドレスについてＦＥＣ受信管理テーブルに登録する（921）。

上記チェック９０２の結果、否であればデスティネーションポート番号がＦＥＣ

通信用か否かがチェックされる（903）。このチェック９０３で、否の場合にはそのパケットをＦＥＣ処理しないでそのまま物理層1111へ送る（922）。前記デスティネーションポート番号のチェック９０３の結果、ＦＥＣ通信用であれば、ＦＥＣ許可テーブル情報の変更パケットか否かがチェックされる（904）。その結果、Yesであれば、ＦＥＣ許可テーブルの内容を更新して（923）、ＴＣＰ処理を行う（924）。即ちＡＣＫを返信して、処理を終わる。

【００４０】

一方、ＴＣＰで無ければ、ＵＤＰパケットか否かがチェックされ(905)、ＵＤＰパケットであれば、ソースアドレスがＦＥＣ受信管理テーブルに在るか否かがチェックされる（906）。ソースアドレスがこの受信管理テーブルに在れば、パケットが破損しているか否かをチェックし（907）、そしてタイムアウトになったＦＥＣシーケンスのパケットか否かがチェックされる（908）。これらのチェックの結果、Yesであれば、ＦＥＣ受信管理テーブルについてソースアドレスからパケット到着率を更新処理し（925）、そのパケットを廃棄して終わる（926）。

一方、前記チェック９０８によりタイムアウトになったＦＥＣシーケンスのパケットでなければ、ＦＥＣ受信管理テーブルについてソースアドレスからパケットの到達率を更新処理する（909）。その後、デコード済みのＦＥＣシーケンスＩＤを持つか否かがチェックされ（910）、Ｙｅｓであれば、そのパケットを破棄して終わる（926）。デコード済みのＦＥＣシーケンスＩＤで無ければ、続いて、そのＦＥＣシーケンスについてＦＥＣデコード用バッファの領域が割り当てられているか否かがチェックされる（911）。その結果否ならば、ＦＥＣシーケンスのデコード処理を開始する（927）。これは例えばＦＥＣデコード用バッファの未使用領域を割り当てて、タイマセットすることにより行われる。

【００４１】

ステップ９１１のチェックで、該当すれば、そのパケットを格納するアドレスを計算して（912）、そのパケットを該当するアドレスへ格納する（913）。

次に、そのパケットがＦＥＣ制御パケットか否かをチェックする（914）。チェックの結果、否であれば、そのパケットがデータパケットか否かをチェックし（

928）、データパケットであれば、そのデータをカプセル化したパケットとして物理層1111に送る（929）。

次にＦＥＣシーケンス（この場合冗長度はｎ）についてそのパケットがｎ個目に届いたか否かをチェックする（915）。その結果該当すれば、続いてＦＥＣシーケンスについて未到着のパケットは冗長パケットか否かがチェックされる（916）。冗長パケットであれば、ＦＥＣ＿ＡＣＫを返信して、シーケンスＩＤを格納する（917）。そしてこの受信ＦＥＣシーケンスの処理を終了する（918）。即ち、ＦＥＣデコード用バッファの割り当てを開放して、シーケンスＩＤを処理済みシーケンスとしてＦＥＣ受信管理テーブルに一時的に格納する動作をする。

【００４２】

上記ステップ９１６における冗長パケットか否かのチェックで、該当しなければ、冗長度ｎ（ステップ８２２により冗長パケットのＦＥＣヘッダが冗長度ｍを示すならば、ｍ）に従って未到着のパケットを再構成する（930）。再構成されたパケットがＦＥＣ制御パケットか否かがチェックされ（931）、該当しなければ、再構成したパケットがＦＥＣデータパケットか否かをチェックする（932）。そしてＦＥＣデータパケットであれば、再構成したデータをカプセル化したパケットとして物理層1111へ送る（934）。尚、ＦＥＣ制御パケットの処理は後述する。

【００４３】

次に図１３乃至図１５を参照して、データの受信時のエラーが発生した場合の処理、及びパケット喪失に係わる処理動作等について説明する。

【００４４】

（Ａ）受信データ不足時の処理

この処理はＦＥＣ冗長化されたＦＥＣシーケンスが一定時間の内に符号化できるほど十分に到達しなかった場合の処理である。

図13に示すフローチャートにおいて、あるシーケンスのパケットの受信を開始すると、タイマ割り込みをする（1501）。そしてあるシーケンス（冗長度ｎ）について、処理を開始してから受信パケットがｎ個未満のままで一定時間が経過したか否かがチェックされる（1502）。ｎ未満であれば、ＦＥＣシーケンスの未到

達パケットの数ｘを算出する（1503）。これは例えば、（ｎ－到達パケット数）の計算をすることにより算出可能である。

次にＦＥＣシーケンスを格納したデコード用バッファ内をクリアして（1504）、パケット到達率の更新を行って、終わる（1505）。即ちこのＦＥＣシーケンスのうち、未到達のパケットは喪失したものとして扱い、パケット到達率を計算する。その計算結果が更新されたパケット到達率となる。

【００４５】

（Ｂ）パケットの到達率の報告処理

算出されたパケット到達率は受信側では受信管理テーブルに登録されると共に、送信側にも送られ、そこでパケット喪失率から冗長度が求められ、冗長度のデータとして送信管理テーブルに登録される。

図１４に示すフローチャートにおいて、受信側ではタイマ割り込みを行い（1601）、各ＦＥＣソースアドレスに対して、パケット到達率を報告するための管理フレームを送信する（1602）。パケット到達率報告のパケットはＦＥＣエンコード用バッファに格納される。パケット到達率（Num_Alive PKT）は、破損、タイムアウトしないで届いたＦＥＣデータパケット数と、送信されたＦＥＣデータパケット数の比を取ることにより算出される（1603）。

【００４６】

（Ｃ）ＦＥＣ制御パケットの処理

図１５に示すフローチャートにおいて、送信側では、届いたパケットがパケット喪失率の報告パケットか否かをチェックする（1701）。喪失率の報告パケットであれば、報告されたパケット喪失率と、予め決められた安全係数から、そのデスティネーション宛ての冗長度を算出して決められる（1702）。求められた冗長度は送信管理テーブルの該当する個所を更新する形で登録される。また、上記ステップ１７０１で、パケット喪失率報告用のパケットで無ければ、管理テーブル変更のパケットか否かがチェックされる（1704）。このチェック１７０４で、該当すれば、送信管理テーブルの内容が更新される（1705）。一方、管理テーブル変更パケットでもなければ、ＦＥＣ＿ＡＣＫか否かがチェックされる（1706）。その結果該当すれば、送信ＦＥＣシーケンスの処理を終了する（1707）。この処理

はＦＥＣエンコード用バッファの割り当て領域を開放し、ＦＥＣ＿ＡＣＫ待ちを解除し、ＦＥＣ送信管理テーブルから削除する処理である。

【００４７】

以上、本発明の実施例を説明したが、上記実施例に限定されずに、種々変形して実施し得る。

【００４８】

図１６はストレージ装置のチャネルアダプタの構成を変形した例である。このチャネルアダプタ１０７に、図２を参照して述べたＦＥＣ変換アダプタの持つ機能を併合したものである。即ち、図２に示した構成のうち、チャネルアダプタ側にある構成、バッファ1102、1112、及び物理層1101、1111が省略され、チャネルアダプタ107はＳＣＳＩコマンド制御部1071、プロトコル制御部1072を有しているが、送信バッファ1073、受信バッファ1074が兼用される。残余の構成は図２と同様である。

【００４９】

また、上記実施例ではパケットの喪失率に従って、送信先毎に冗長度を変更してデータを送信するようにしたが、パケットの喪失率に従ってパケット送信の間隔を変えることも可能である。広義にはパケット喪失率に従ってデータ送信の状態を変えるようにできる。

【００５０】

【発明の効果】

本発明によれば、FEC方式を採用したiSCSIプロトコル準拠のストレージ装置を提供することができる。また、受信側で把握したパケットの喪失状況を送信側にフィードバックすることにより、送信側ではデータの送信先毎にFEC処理の状態やデータ送信の冗長度を変えてデータ送信することができ、一方受信側ではそれに応じて復元することができる。これによりパケットが喪失した場合でもデータの復元が可能はストレージシステムを得ることができる。

【図面の簡単な説明】

【図１】

本発明の一実施例によるストレージシステムを示すブロック図。

【図２】

本発明の一実施例に係るＦＥＣ変換アダプタを有するストレージ装置の一例を示す図。

【図３】

本発明の他の実施例によるストレージシステムを示すブロック図。

【図４】

本発明の一実施例によるデータ転送の概念を示す図。

【図５】

通信に用いられるパケットのフォーマットの一例を示す図。

【図６】

ＦＥＣヘッダのフォーマットを示す図。

【図７】

ＦＥＣの制御テーブルと送信及び受信管理テーブルの内容の一例を示す図。

【図８】

データを送信するときのＦＥＣエンコード処理の動作を説明するためのフローチャート。

【図９】

ＦＥＣエンコード処理における送信データ不足時の動作を説明するためのフローチャート。

【図１０】

ＦＥＣパケットの再送処理動作を説明するためのフローチャート。

【図１１】

データ受信時におけるＦＥＣデコード処理の動作を説明するためのフローチャート。

【図１２】

データ受信時におけるＦＥＣデコード処理の動作を説明するためのフローチャート。

【図１３】

データの受信時にエラーが発生した場合の処理動作を説明するためのフローチャ

ート。

【図１４】

データの受信時におけるパケット到達率の報告処理動作を説明するためのフローチャート。

【図１５】

パケット喪失の報告に対する処理動作について説明するためのフローチャート。

【図１６】

本発明の他の実施例によるストレージ装置の構成を示す図。

【符号の説明】

100,200:ストレージ装置　　300:ホストコンピュータ

400:ネットワーク　　500:管理サーバ

101,201:ディスクドライブ　　103,203:ディスクアダプタ

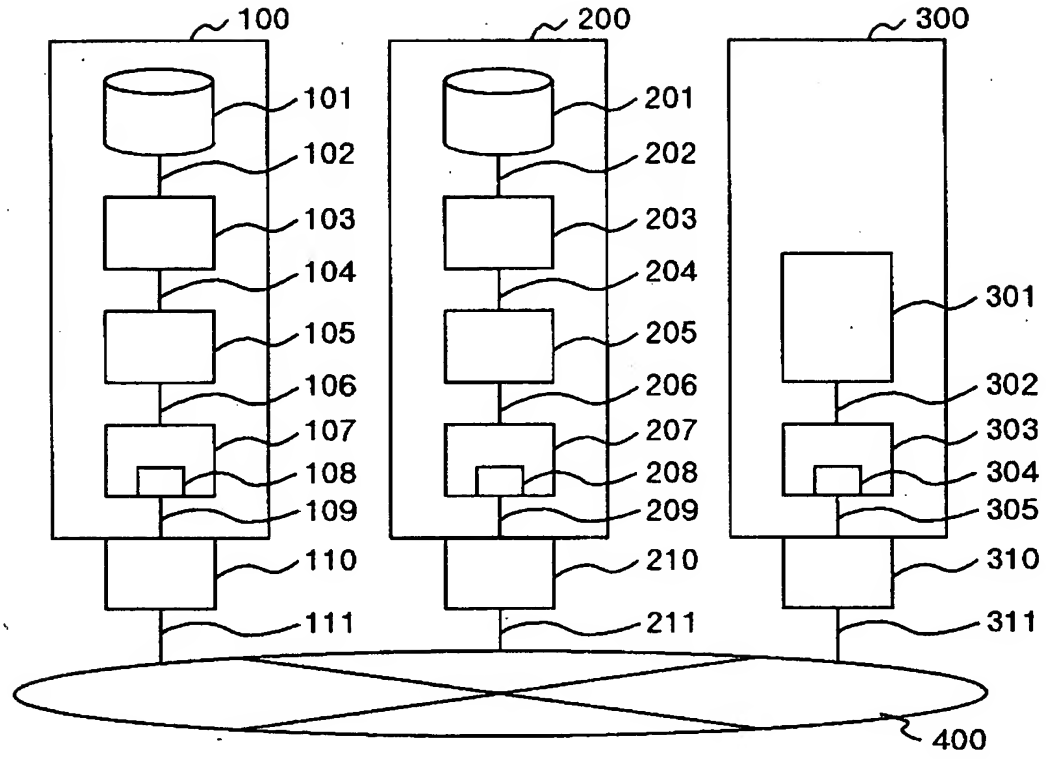
105,205:キャッシュメモリ　　107,207:チャネルアダプタ

110,210,310:ＦＥＣ変換アダプタ

【書類名】　図面

【図１】

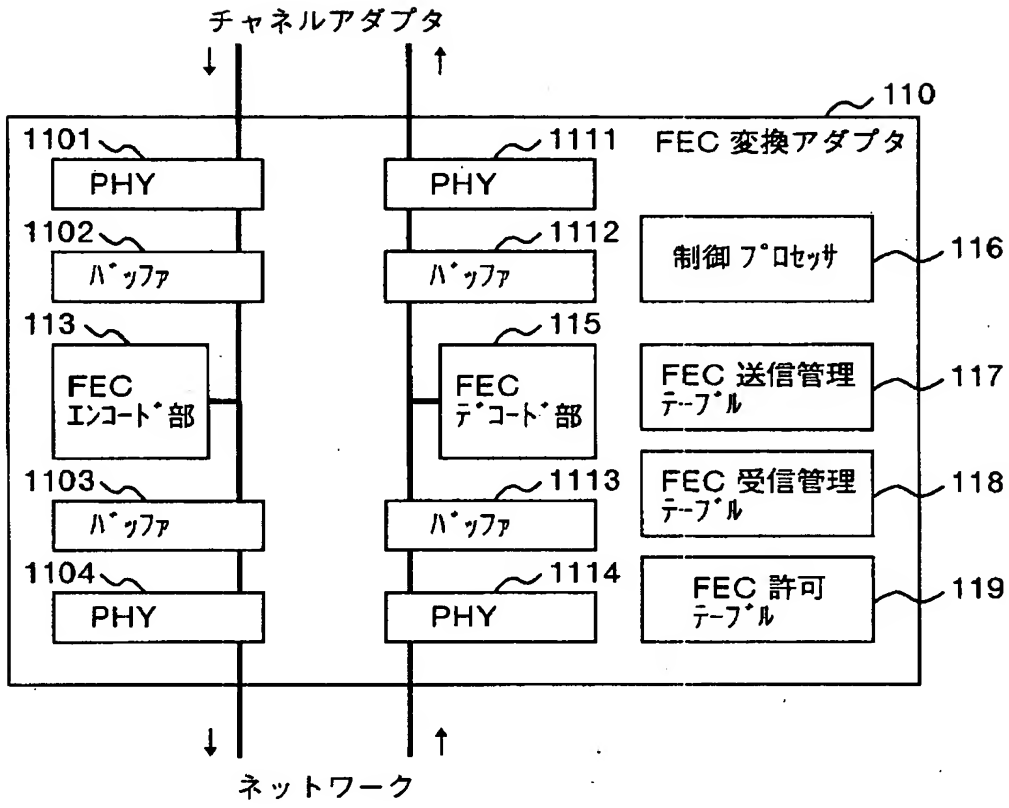
図　1



出証特２００３－３０４１８９７

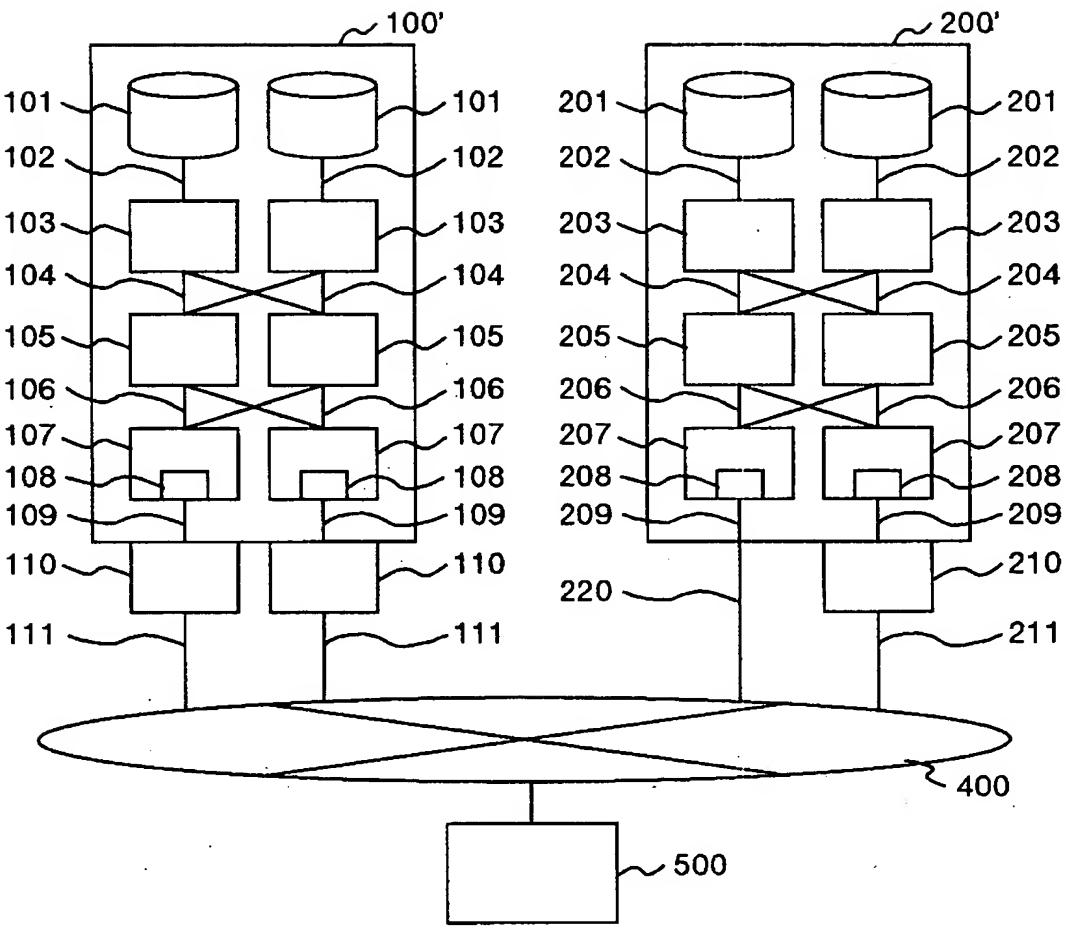
【図２】

図　2


チャネルアダプタ
110
FEC 変換アダプタ
1101
PHY
1111
PHY
1102
ﾊﾞｯﾌｧ
1112
ﾊﾞｯﾌｧ
制御 ﾌﾟﾛｾｯｻ
116
113
FEC
ｴﾝｺｰﾄﾞ部
115
FEC
ﾃﾞｺｰﾄﾞ部
FEC 送信管理
ﾃｰﾌﾞﾙ
117
1103
ﾊﾞｯﾌｧ
1113
ﾊﾞｯﾌｧ
FEC 受信管理
ﾃｰﾌﾞﾙ
118
1104
PHY
1114
PHY
FEC 許可
ﾃｰﾌﾞﾙ
119
ネットワーク

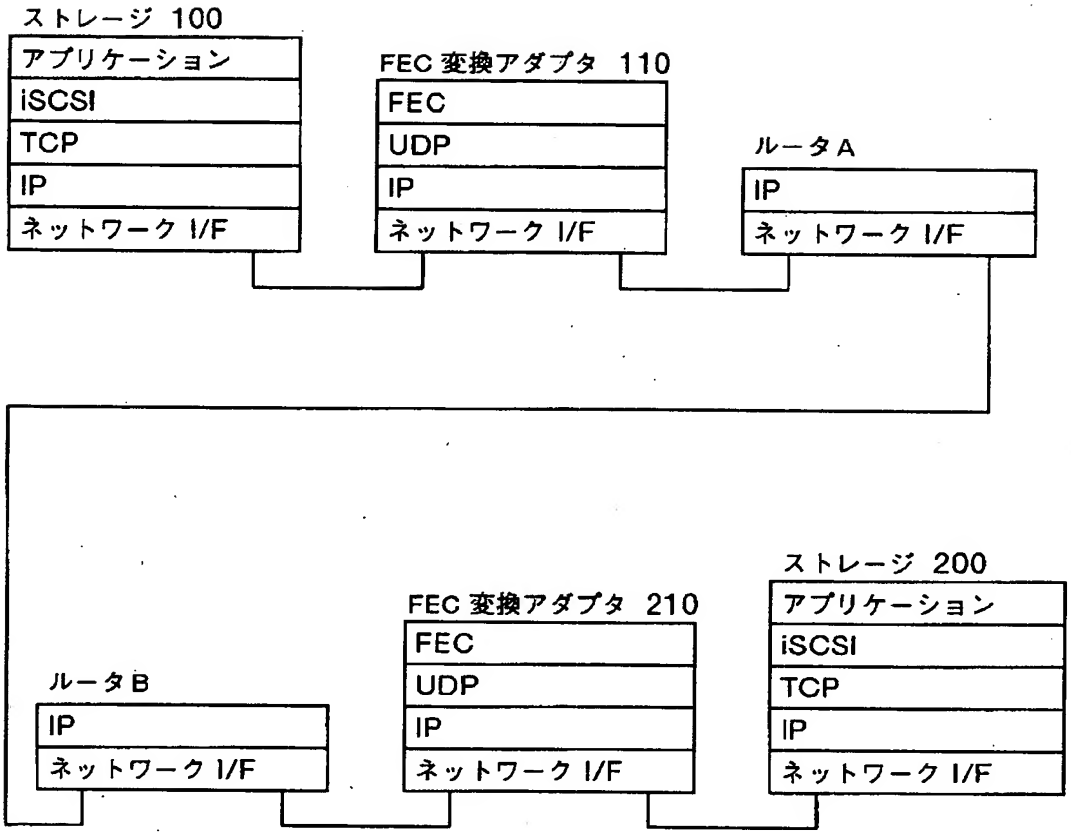
【図３】

図　3


100'
200'
101
102
103
104
105
106
107
108
109
110
111
201
202
203
204
205
206
207
208
209
210
211
220
400
500

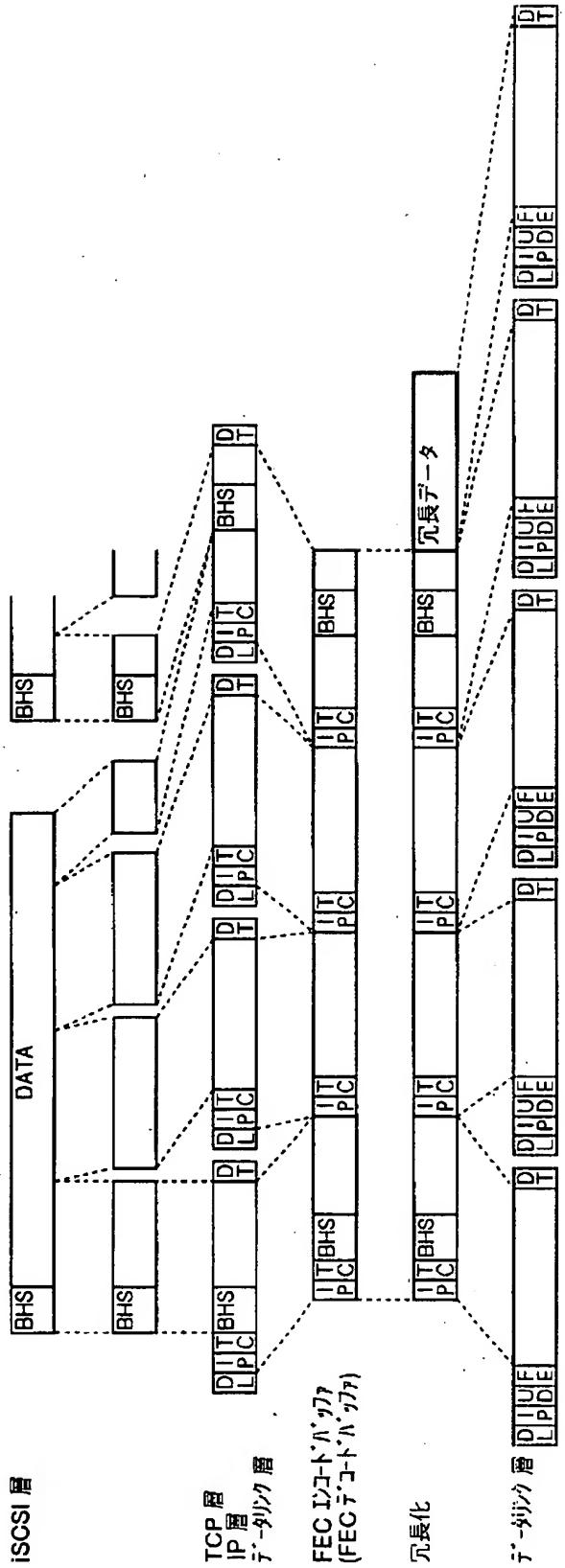
【図４】

図　4


ストレージ 100
アプリケーション
iSCSI
TCP
IP
ネットワーク I/F
FEC 変換アダプタ 110
FEC
UDP
IP
ネットワーク I/F
ルータA
IP
ネットワーク I/F
ストレージ 200
アプリケーション
iSCSI
TCP
IP
ネットワーク I/F
FEC 変換アダプタ 210
FEC
UDP
IP
ネットワーク I/F
ルータB
IP
ネットワーク I/F

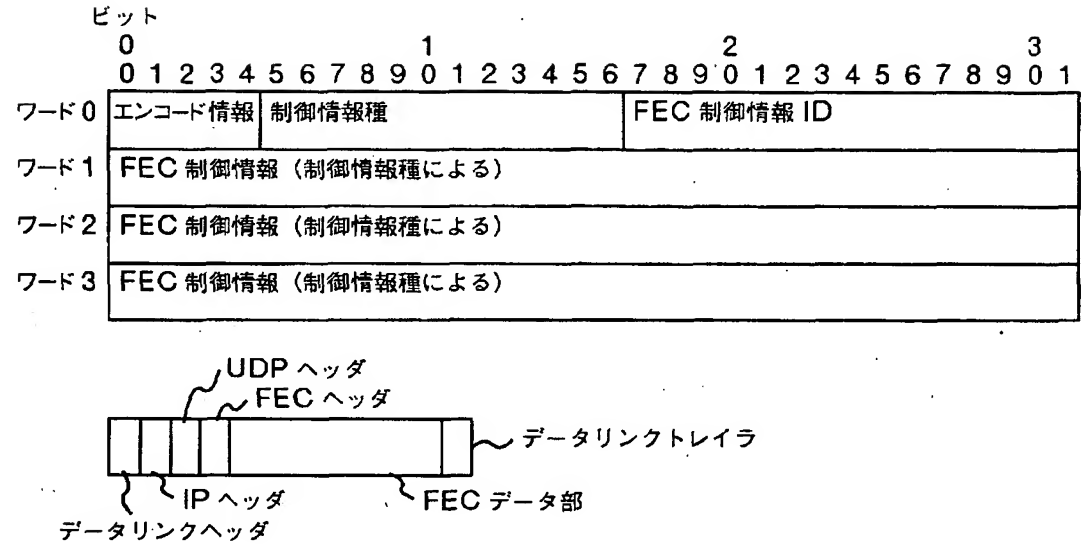
【図５】

図 5


iSCSI 層
TCP 層
IP 層
データリンク 層
FEC エンコードバッファ
(FECデコードバッファ)
冗長化
データリンク 層
BHS
DATA
冗長データ
DL : データリンクヘッダ
IP : IP ヘッダ
TC : TCP ヘッダ
DT : データリンクトレイラ
UD : UDP ヘッダ
FE : FEC ヘッダ

【図６】

図　6


ビット
0
1
2
3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
ワード0
エンコード情報
制御情報種
FEC 制御情報 ID
ワード1
FEC 制御情報（制御情報種による）
ワード2
FEC 制御情報（制御情報種による）
ワード3
FEC 制御情報（制御情報種による）
UDP ヘッダ
FEC ヘッダ
データリンクトレイラ
IP ヘッダ
FEC データ部
データリンクヘッダ

【図７】

図　7

FEC 許可テーブル 119

通信相手の iSCSI Name	初期冗長度 (n)

FEC 送信管理テーブル 117

デスティネーションアドレス	冗長度 (n)	FEC エンコード用バッファ制御情報

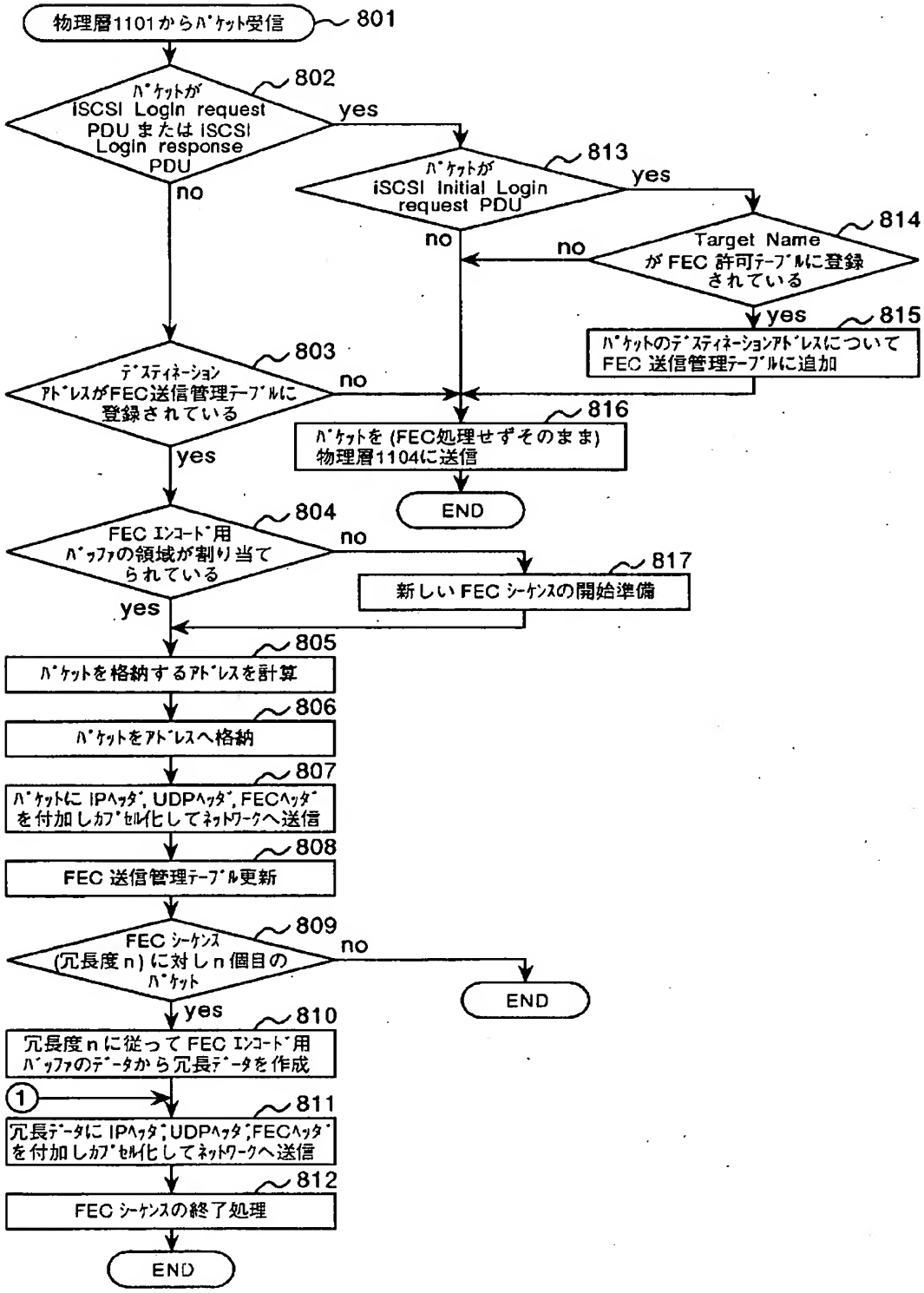
FEC 受信管理テーブル 118

ソースアドレス	パケット到達率	FEC デコード用バッファ制御情報
	／	
	／	
	／	
	／	

【図８】

図　8

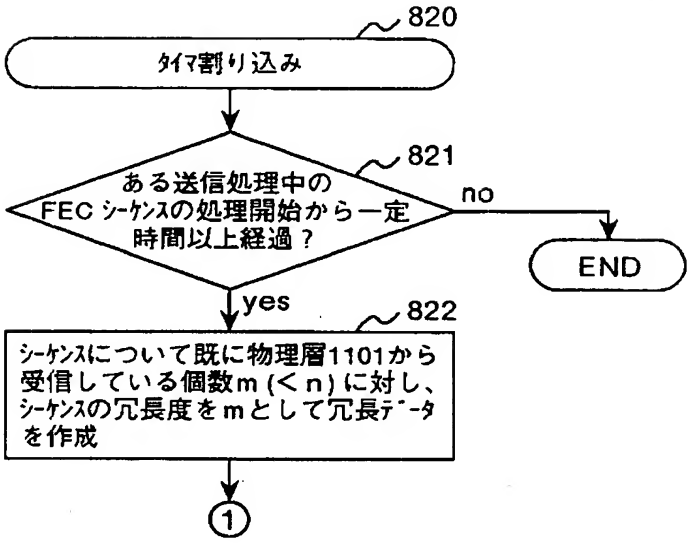
エンコード処理



【図９】

図　９

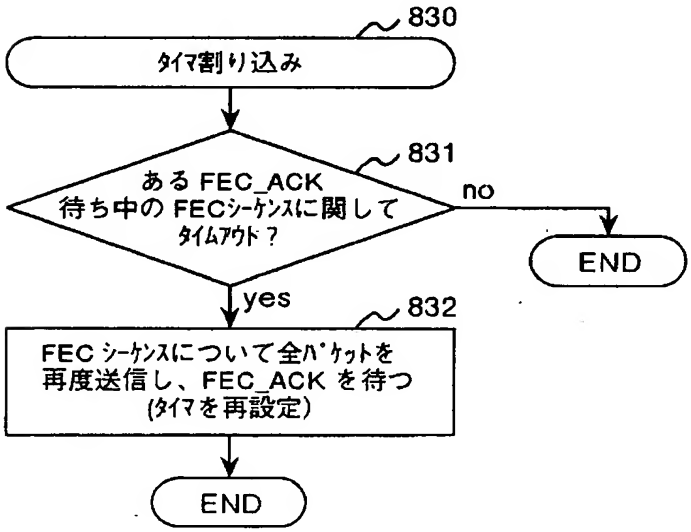
送信データ不足時の処理



【図１０】

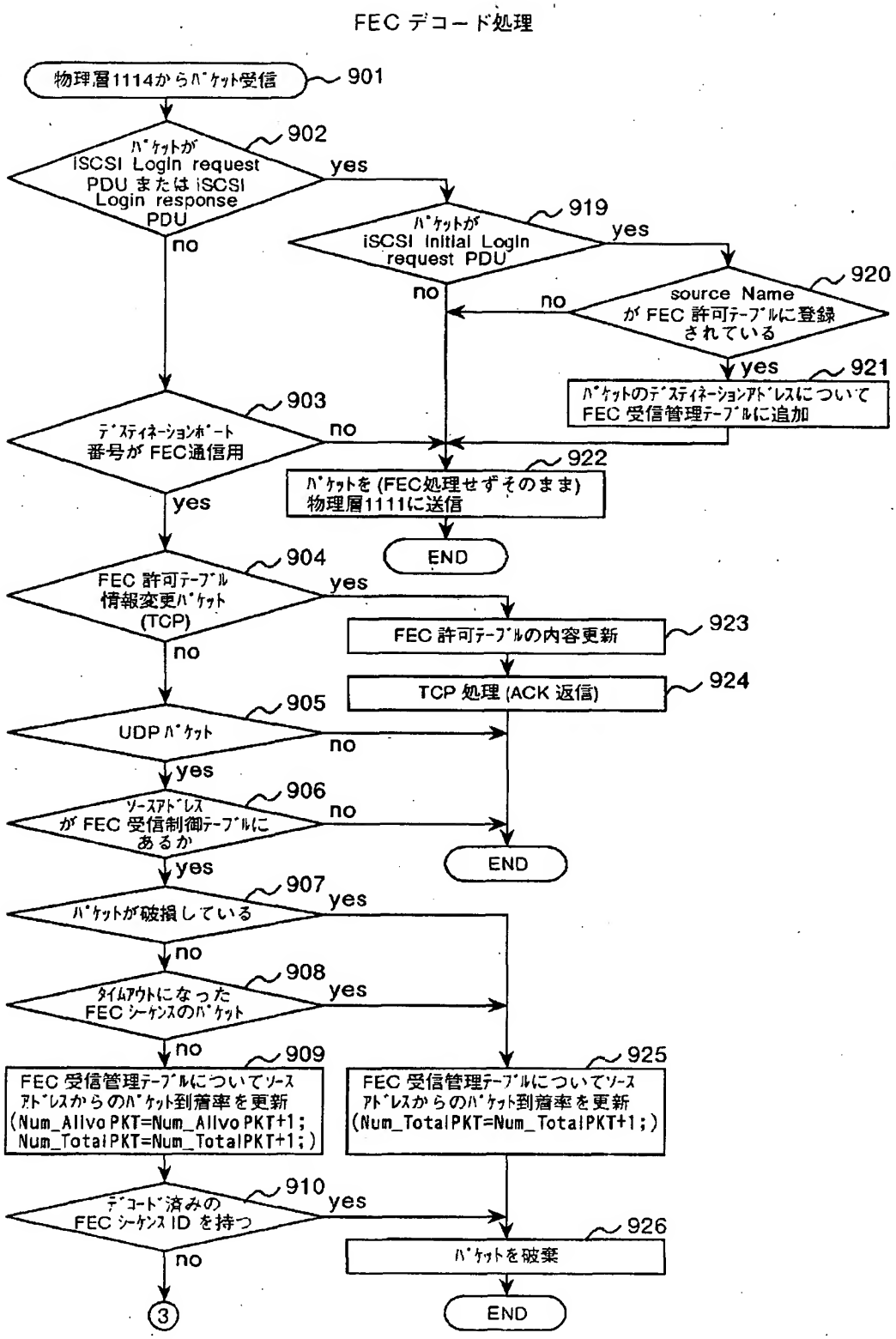
図　１０

FEC_ACK 待ちタイムアウトによる再送処理



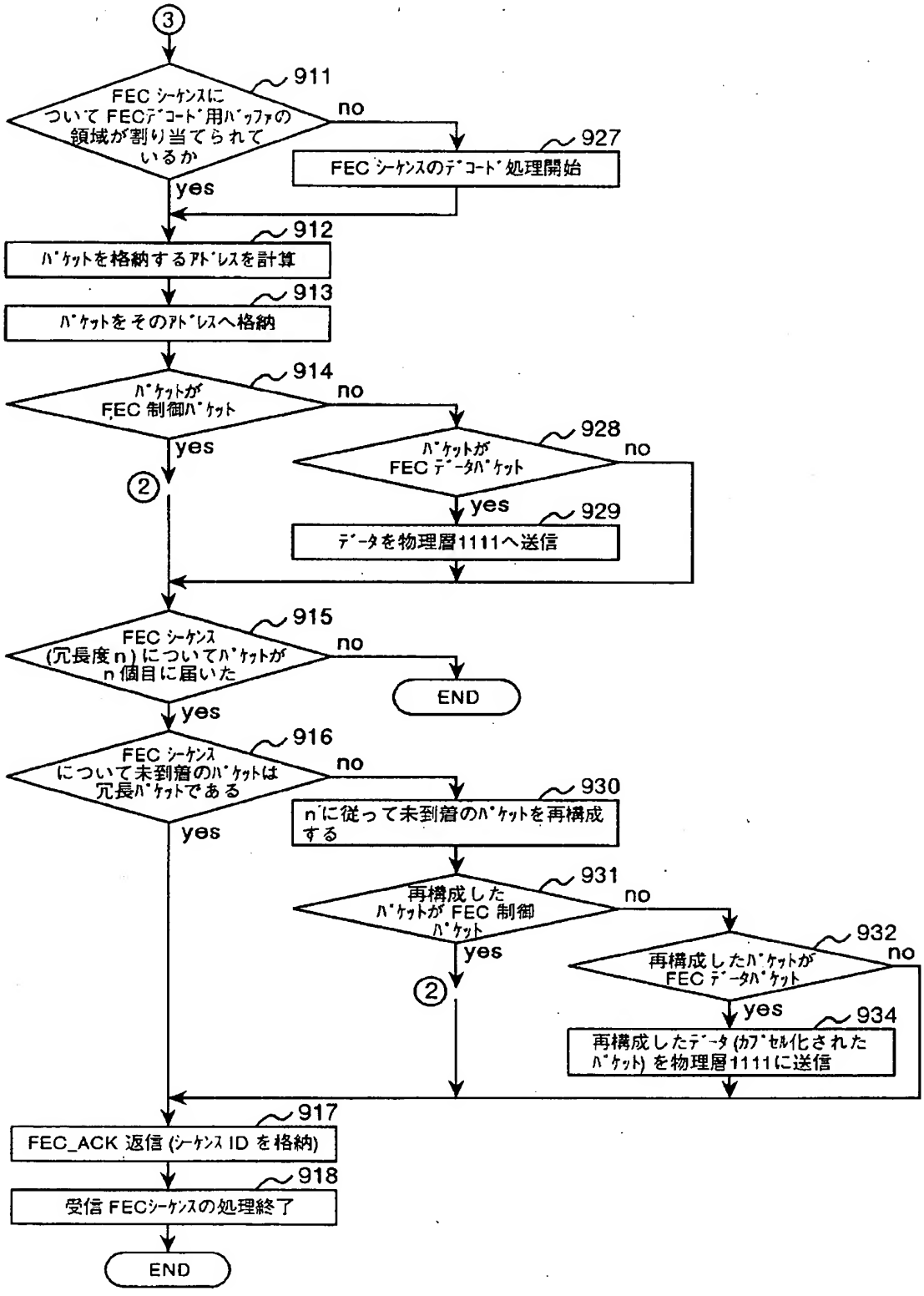
【図１１】

図　１１


FEC デコード処理
物理層1114からﾊﾟｹｯﾄ受信
901
ﾊﾟｹｯﾄが iSCSI Login request PDU または iSCSI Login response PDU
902
yes
no
ﾊﾟｹｯﾄが iSCSI Initial Login request PDU
919
yes
no
source Name が FEC 許可ﾃｰﾌﾞﾙに登録されている
920
no
yes
ﾊﾟｹｯﾄのﾃﾞｽﾃｨﾈｰｼｮﾝｱﾄﾞﾚｽについて FEC 受信管理ﾃｰﾌﾞﾙに追加
921
ﾃﾞｽﾃｨﾈｰｼｮﾝﾎﾟｰﾄ番号が FEC通信用
903
no
yes
ﾊﾟｹｯﾄを (FEC処理せずそのまま) 物理層1111に送信
922
END
FEC 許可ﾃｰﾌﾞﾙ情報変更ﾊﾟｹｯﾄ (TCP)
904
yes
no
FEC 許可ﾃｰﾌﾞﾙの内容更新
923
TCP 処理 (ACK 返信)
924
UDP ﾊﾟｹｯﾄ
905
no
yes
ｿｰｽｱﾄﾞﾚｽが FEC 受信制御ﾃｰﾌﾞﾙにあるか
906
no
yes
END
ﾊﾟｹｯﾄが破損している
907
yes
no
ﾀｲﾑｱｳﾄになった FEC ｼｰｹﾝｽのﾊﾟｹｯﾄ
908
yes
no
FEC 受信管理ﾃｰﾌﾞﾙについてｿｰｽｱﾄﾞﾚｽからのﾊﾟｹｯﾄ到着率を更新 (Num_AllvoPKT=Num_AllvoPKT+1; Num_TotalPKT=Num_TotalPKT+1;)
909
FEC 受信管理ﾃｰﾌﾞﾙについてｿｰｽｱﾄﾞﾚｽからのﾊﾟｹｯﾄ到着率を更新 (Num_TotalPKT=Num_TotalPKT+1;)
925
ﾃﾞｺｰﾄﾞ済みの FEC ｼｰｹﾝｽ ID を持つ
910
yes
no
ﾊﾟｹｯﾄを破棄
926
END
③

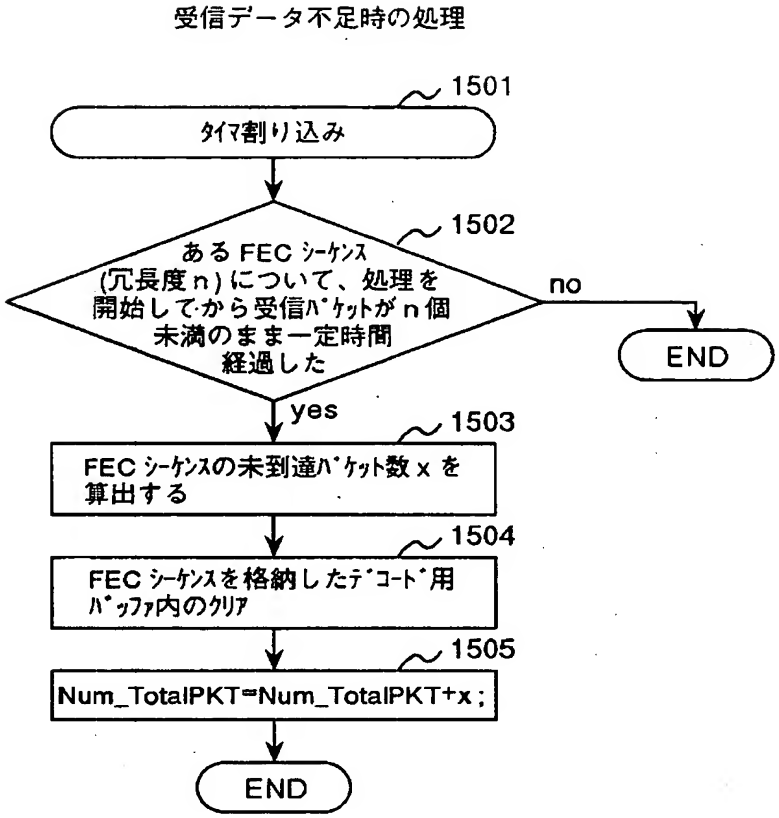
【図１２】

図　１２


③
911
FEC シーケンスについてFECデコード用バッファの領域が割り当てられているか
no
927
FEC シーケンスのデコード処理開始
yes
912
パケットを格納するアドレスを計算
913
パケットをそのアドレスへ格納
914
パケットがFEC 制御パケット
no
928
パケットがFEC データパケット
no
yes
929
データを物理層1111へ送信
②
915
FEC シーケンス(冗長度ｎ)についてパケットがｎ個目に届いた
no
END
yes
916
FEC シーケンスについて未到着のパケットは冗長パケットである
no
930
ｎに従って未到着のパケットを再構成する
yes
931
再構成したパケットがFEC 制御パケット
no
932
再構成したパケットがFEC データパケット
no
yes
②
934
再構成したデータ(カプセル化されたパケット)を物理層1111に送信
917
FEC_ACK 返信(シーケンス ID を格納)
918
受信 FECシーケンスの処理終了
END

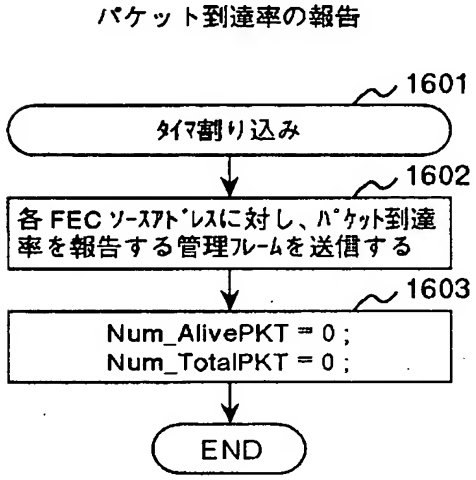
【図１３】

図　１３



【図１４】

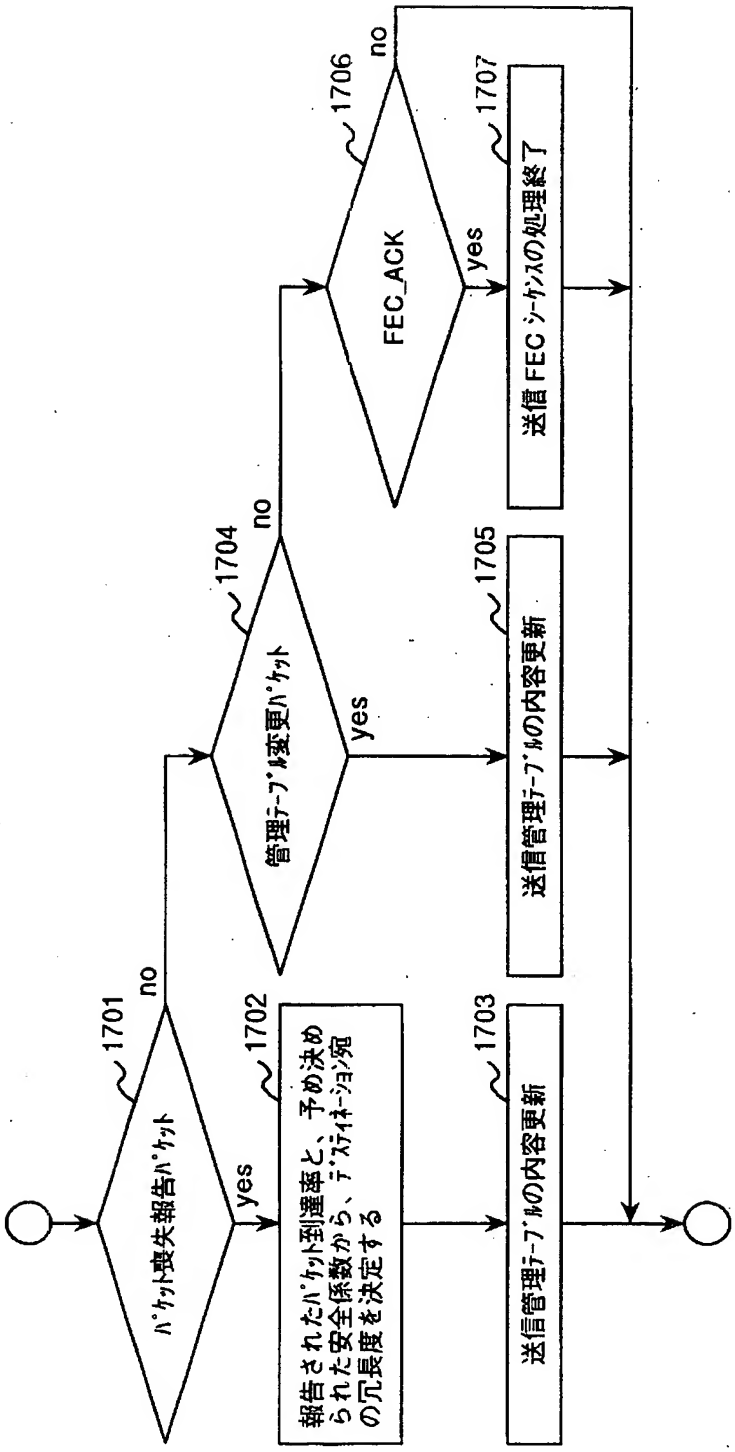
図　１４



【図１５】

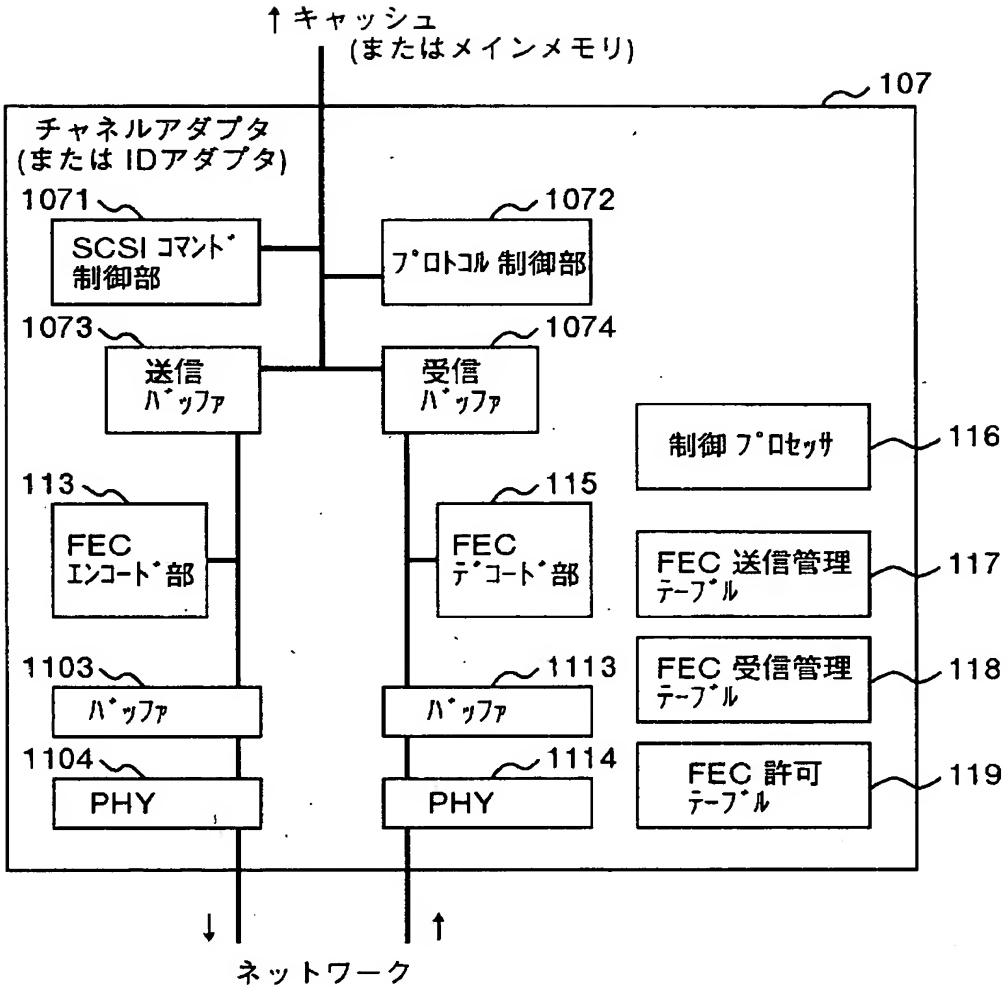
図　１５

FEC 制御パケット処理



【図１６】

図　１６


↑キャッシュ
(またはメインメモリ)
107
チャネルアダプタ
(または IDアダプタ)
1071
SCSI ｺﾏﾝﾄﾞ
制御部
1072
ﾌﾟﾛﾄｺﾙ 制御部
1073
送信
ﾊﾞｯﾌｧ
1074
受信
ﾊﾞｯﾌｧ
制御 ﾌﾟﾛｾｯｻ
116
113
FEC
ｴﾝｺｰﾄﾞ部
115
FEC
ﾃﾞｺｰﾄﾞ部
FEC 送信管理
ﾃｰﾌﾞﾙ
117
1103
ﾊﾞｯﾌｧ
1113
ﾊﾞｯﾌｧ
FEC 受信管理
ﾃｰﾌﾞﾙ
118
1104
PHY
1114
PHY
FEC 許可
ﾃｰﾌﾞﾙ
119
↓
↑
ネットワーク

【書類名】　要約書

【要約】

【課題】

FEC方式を採用したiSCSIプロトコル準拠のストレージ装置間のデータ送受信において、パケットが喪失した場合でも送信データの復元が可能なストレージシステムを実現する。

【解決手段】

iSCSIプロトコル準拠のストレージ装置がネットワークを介して接続されたストレージシステムにおいて、ストレージ装置にFECの機能を持たせ、iSCSI層間でデータの伝送を行う場合、データの送信先毎にFEC処理の状態やデータ送信の冗長度を変えてデータ送信し、復元する。

例えば、ネットワーク側のポートとストレージ装置側のポートとの間に介在し、パケットデータの送受信を行うシステムにおいて、送信先に対応してＦＥＣ冗長度を登録して管理する送信管理テーブルと、受信先に対応してＦＥＣ冗長度を登録して管理する受信管理テーブルと、ストレージ装置で生成されるパケット化されたiSCSI層のデータを、該送信管理テーブルを参照して送信先に応じた冗長度を持たせてＦＥＣエンコード処理するエンコード部と、該受信管理テーブルを参照して、ネットワークから受信されるパケットデータをＦＥＣデコード処理し、iSCSI層のデータに復号化するデコード部とを有する。

【選択図】　図１

出証特２００３－３０４１８９７

出　願　人　履　歴　情　報

識別番号　　〔０００００５１０８〕

１．変更年月日　　１９９０年　８月３１日
［変更理由］　新規登録
住　所　東京都千代田区神田駿河台４丁目６番地
氏　名　株式会社日立製作所